VERİ MADENCİLİGİ RAPOR

Proje Konusu : Car\_Prices\_Poland.csv ,audi.csv ,bmw.csv ,skoda.csv , toyota.csv , car\_ad.csv deki verileri kullanarak bazı araba markalarının (BMW , AUDİ , SKODA, TOYOTA ) fiyatının tahmini yapmak için veri ambarı hazırlanması ve yorumlanması.

Veri Setleri Linkleri

**Car\_ad.csv :** <https://www.kaggle.com/datasets/antfarol/car-sale-advertisements?select=car_ad.csv>

**Skoda.csv :** <https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes?select=skoda.csv>

**Bmw.csv :** <https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes?select=bmw.csv>

**Toyota.csv :** <https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes?select=toyota.csv>

**Audi.csv :** <https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes?select=audi.csv>

**Car\_Prices\_Poland\_Kaggle.csv :** https://www.kaggle.com/datasets/aleksandrglotov/car-prices-poland

Veri Setleri Hakkında Bilgilendirme Ve Hazırlanması

**Car\_ad.csv deki Verilerin Sütunlari**

* Car : Arabanın markası ( Audi , Skoda , Toyota, BMW, Diğer Markalar )
* Price : Arabanın fiyatı ( para birimi USD )
* Mileage : Ne kadar yol gittiği (KM)
* engV : Motor Hacmi (0 – 99 arası )
* Year : Üretim Yılı
* EngyType : Yakıt türü (Petrol ,Diesel ,Gas,Gassoline,CNG,Electric,Hybrid,Diğerleri )

**Car\_ad.csv deki Verilerin Düzenlenmesi**

1. Veri setindeki kullanılmayan sütünlar çıkartıldı.
2. Audi,Skoda,Toyota,BMW markaları filtrelendi
3. Fiyat bilgisi Euro ya çevrildi.
4. Null değere sahip satırlar silindi.
5. Car = Marka,year = ÜretimYılı, mileage=KM, engV= MotorHacmi, engType=YakıtTipi , price=Fiyatı şeklinde sütunların isimleri değiştirildi.

**Audi.csv ,Bmw.csv , Skoda.csv, Toyota.csv deki Verilerin Sütunları**

* Model : Arabanın markası
* Price : Arabanı fiyatı (para birimi Euro )
* Mileage : Ne kadar yol gittiği (Mil)
* engineSize : Motor Hacmi (0 – 99 arası )
* Year : Üretim Yılı
* fuelType : Yakıt türü (Petrol ,Diesel )

**Audi.csv ,Bmw.csv , Skoda.csv, Toyota.csv deki Verilerin Düzenlenmesi**

1. Veri setindeki kullanılmayan sütünlar çıkartıldı.
2. Ne kadar yol gittiği mil den km ye çevrildi.
3. Null değere sahip satırlar silindi.
4. Marka =Marka, year=ÜretimYılı, mileage=KM, engineSize=MotorHacmi, fuelType=YakıtTipi, price=Fiyatı şeklinde sütun isimleri değiştirildi

**Car\_Prices\_Poland\_Kaggle.csv deki Verilerin Sütunları**

* Mark : Arabanın markası (Audi , Diğer Markalar)
* Price : Arabanın fiyatı (para birimi Polonya Lirası )
* Mileage : Ne kadar yol gittiği (KM)
* Vol\_engine : Motor hacmi (0-7600 arası )
* Year : Üretim Yılı
* Fuel : Yakıt türü (Diesel , Petrol ,LPG , Hybrid ,Gasoline,Gas)

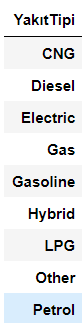
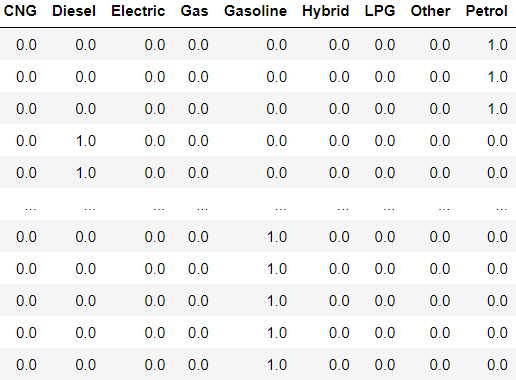
**Car\_Prices\_Poland\_Kaggle.csv deki Verilerin Düzenlenmesi**

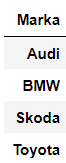
1. Veri setindeki kullanılmayan sütünlar çıkartıldı.
2. Veri setinden Audi markalı araçlar filtrelendi.
3. Araba fiyatı Polonya Lirasından Euro ya çevrildi.
4. Motor Hacmi 1000 bölünüp yuvarlanarak diğer motor hacimleriyle aynı türde oldu 7600 => 7,6 gibi.
5. Null değere sahip satırlar silindi.
6. Mark=Marka, year=ÜretimYılı, mileage=KM, vol\_engine=MotorHacmi, fuel=YakıtTipi, price=Fiyatı şeklinde sütun isimleri düzenlendi.

Veri Ambarını İnceleme

Oluşturulan bütün veri setleri birleştirildi ismi TumVeriler olarak geçiyior.

Oluşan veri ambarındaki kategorik veriler (YakitTipi ve Marka ) sayısal veriye çevrildi.

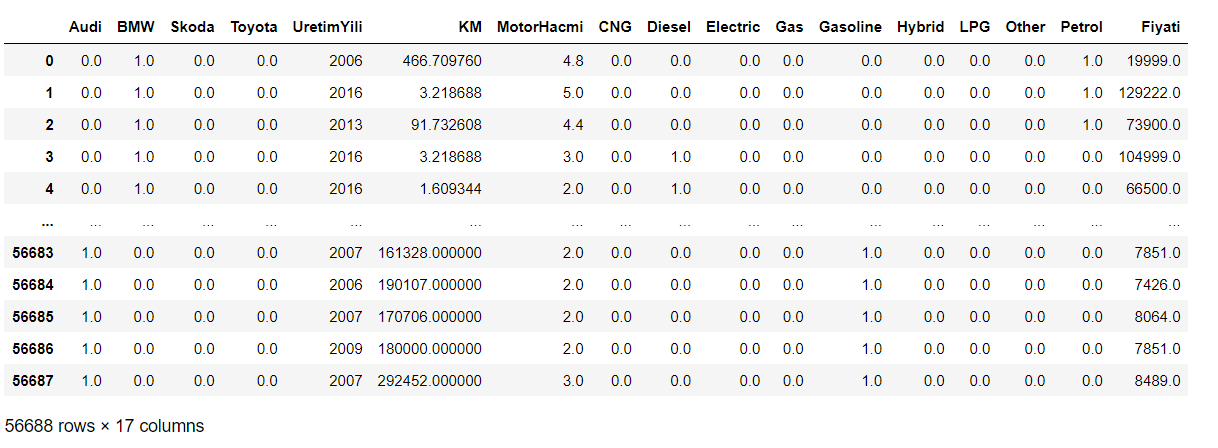
 

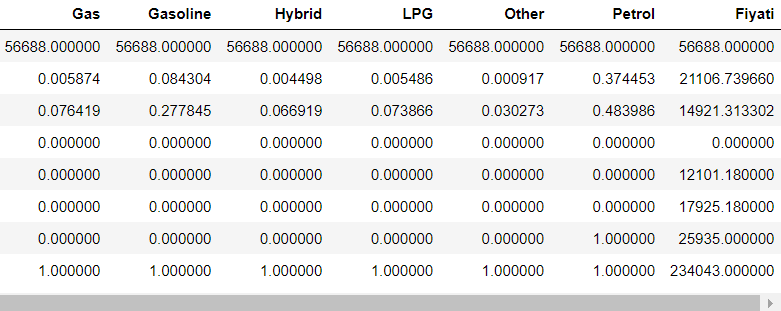
Yakıt türleri verisi hangi değerse mesela petrol ise sadece o sütun değeri 1.0 şeklinde olmuştur.

Marka çeşitleri hangi markaya aitse o değer 1.0 olmuştur.

**Oluşan Yeni Veri Ambarı :**

****

**Veri Ambarının İncelenmesi**



Marka : Toplamda 4 marka vardır. En çok bulunan marka ise Audi markasıdır.

Üretim Yılı : 1979 ile 2022 arasında değişmekte . Veri ambarındaki üretim yılının büyük çoğunluğu yakın bir tarihte 2019.

KM : 0 ile 550000 arasında değişmekte. Veri ambarındaki araçların çoğunun km bilgisi 53224 km civarinda.

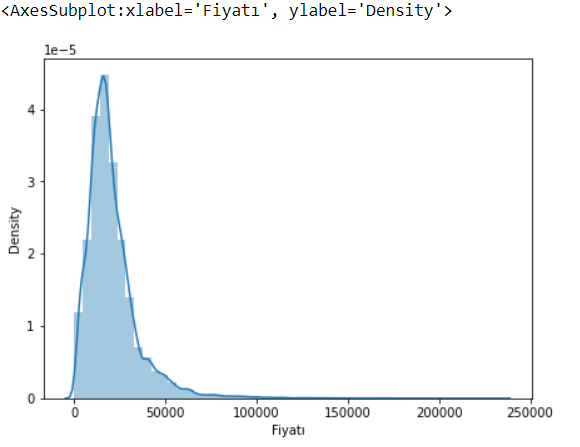
MotorHacmi : 0 ile 99 arasında değişmekte. Veri ambarındaki araçların çoğuda 2.0 lık bir motor hacmine sahip.

YakıtTipi : CNG, Diesel, Electric, Gas, Gasoline, Hybrid, LPG, Petrol ve Other şeklinde toplamda 9 tane yakıt tipi vardır.Arabalarda ise çoğunlukla Petrol ve Diesel yakıt olarak kullanılmıştır.

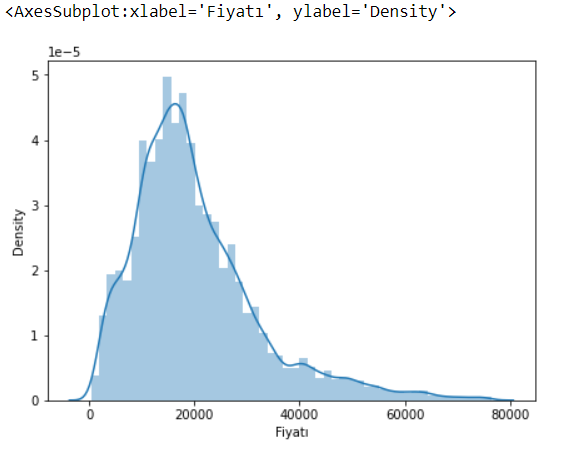
Fiyatı : Veri ambarındaki araçların fiyatı 0 ile 234043 euro arasında değişmektedir ama 0 olan veriler veri ambarından çıkartılmıştır.Araçların ortalama fiyatı ise 20919 euro bandındadır.En düşük araç fiyatı 309 euro iken en yüksek fiyat ise 234043 euro dur.

MotorHacmi ve Fiyatı 0 olamayacağı için o satırlar veri ambarından çıkartıldı.

**Fiyatın Dağılımı :**

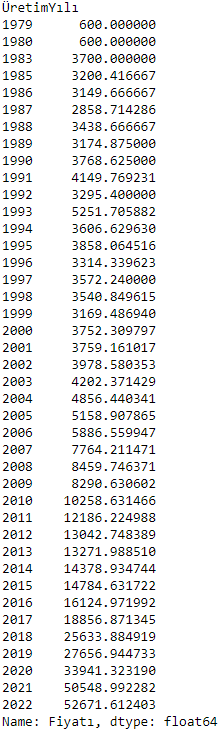


Fiyatın grafiği normal dağılıma sahip fakat belirli bir yerden sonra fiyatın düz bir çizgide hareket etmesi oluşturulan modelin fiyat tahminin olması gerekenden yüksek olmasına sebeb olabilir.Bu yüzden veri ambarından en yüksek fiyata sahip araçların bir kısmını çıkartılablir.Veri ambarımızda toplam 56192 veri var bu verilerin en yüksek fiyata sahip olan %1 lik kısmı veri ambarından çıkartılınca oluşan grafik şu şekilde :

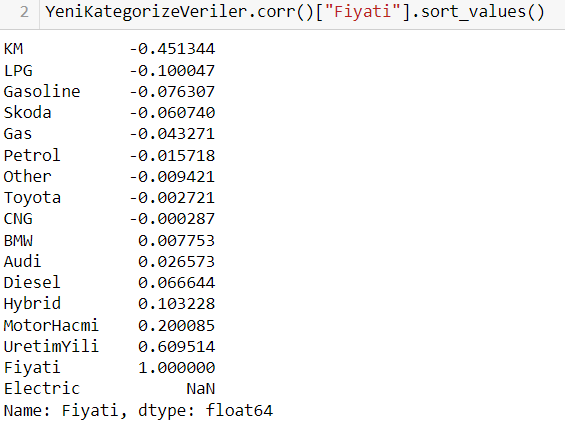


Grafikte görüldüğü gibi eskiye göre daha iyi bir normal dağılıma sahiptir.

Aşadaki Tabloda fiyatların yıllara göre daha orantılı bir şekilde olduğu gözüküyüor :



Veri Ambarının Korelasyon Değerleri



Korelasyon değerleri iki değişken arasındaki ilişkiyi verir.

Değişkenlerin fiyatı etkisi şekildeki gibidir.

* KM arttıkça aracın fiyatı düşüyüor.
* ÜretimYılı büyüdükçe ise aracın fiyatı artıyıor.
* MotorHacmi büyükdüçe ise aracın fiyatı artıyıor.
* Aracın yakıt türü Hybrid , Diesel ise aracın fiyatı artıyıor.Hybrid ise Diesele göre daha pahalı olduğu gözüküyüor.
* Aracın yakıt türü Gasoline, Gas, Petrol, Other, CNG ise arabanın fiyatını düşürüyüorlar.Bunların fiyatı düşürme katsayı ise sırasıyla Gasoline > Gas > Petrol > Other > CNG şeklinde.
* Yakıt türü olan Electric de ise Nan değeri gözüküyüor bunun sebebi ise Fiyat ve Electric arasında korelasyon ilişkisi bulunamamış olmasından dolayı.
* Araba markalarında ise Audi ve BMW markalı araçlar fiyatı yükseliyior.Toyota ve Skoda ise bunlara göre fiyatı düşürüyüor. Yani markaların değerleri arasında Audi > BMW > Toyota > Skoda şeklinde fiyat etkileri vardır.

Hüseyin Oral

170101033